DOCUMENT RESUME

ED 385 596 TM 024 044

AUTHOR	Gallagher, Ann; Mandinach, Ellen
TITLE	Strategy Use on Multiple-Choice and Free-Response Items: An Analysis of Sex Differences among High Scoring Examinees on the SAT-M. Final Report.
INSTITUTION	Educational Testing Service, Princeton, N.J.
REPORT NO	ETS-RR-92-54
PUB DATE	Sep 92
NOTE	53p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Algorithms; *College Students; Higher Education; *Mathematics Tests; *Multiple Choice Tests; *Problem Solving; *Protocol Analysis; Scores; Sex Differences
IDENTIFIERS	*Free Response Test Items; Scholastic Aptitude Test; *Strategy Choice

ABSTRACT

Twenty-four students who scored 650 or more on the Scholastic Aptitude Test Mathematics test (SAT-M) were asked to think aloud while solving 13 mathematics items in either multiple-choice or free-response format. Strategies students used to solve the items were classified as either algorithmic or insightful. Data analyses indicated that items in the free-response format were significantly more difficult for females than multiple choice items. No significant difference was found for males. Females were more likely than males to use algorithmic strategies in both the multiple-choice and free-response formats, with no significant difference in types of strategies used in either format. Males were more likely to use insightful strategies in the multiple-choice format than in the free-response format. Finally, on multiple-choice items, both males and females used options to detect calculation errors, with females using this strategy somewhat more than males. One table gives score distribution by sex. Appendix A contains multiple-choice and free-response versions of items used in the protocols. Appendix B provides samples of solution strategies. (Contains 31 references.) (Author/SLD)

RR-92-54

RESEARCH REPORT

STRATEGY USE ON MULTIPLE-CHOICE AND FREE-RESPONSE ITEMS: AN ANALYSIS OF SEX DIFFERENCES AMONG HIGH SCORING EXAMINEES ON THE SAT-M

Ann Gallagher

with assistance from
Ellen Mandinach

Educational Testing Service
Princeton, New Jersey
September 1992

Strategy Use on Multiple-Choice and Free-Response Items: An Analysis of Sex Differences Among High Scoring Examinees on the SAT-M

Ann Gallagher

with assistance from
Ellen Mandinach

Educational Testing Service

Final Report
Submitted to the Program Research Planning Council
December 1991

ABSTRACT

Twenty-four students who scored 650 or more on the SAT-M were asked to think aloud while solving 13 mathematics items in either multiple-choice or free-response format. Strategies students used to solve the items were classified as either algorithmic or insightful. Data analyses indicated that items in the free-response format were significantly more difficult for females than multiple choice items. No significant difference was found for males. Females were more likely than males to use algorithmic strategies in both the multiple-choice and free-response formats, with no significant difference in types of strategies used in either format. Males were more likely to use insightful strategies in the multiple-choice format than in the free-response format. Finally, on multiple-choice items, both males and females used options to detect calculation errors, however with females using this strategy somewhat more than males.

Introduction

Many studies have explored sex differences in mathematical performance. Maccoby and Jacklin (1974) and Hyde (1981) performed meta-analytic syntheses of this body of work and concluded that sex differences on quantitative measures follow a developmental pattern. Up to about fifth grade few sex differences in mathematical performance are found. Differences favoring males begin to emerge with adolescence. Although Hyde (1981) reports that the differences are small, they are found fairly consistently in studies of adolescent and post-adolescent populations.

Recently, Hyde, Fennema, and Lamon (1990) performed a meta-analysis of a broader group of studies. Although they discovered negligible sex differences in the general population, they did find differences favoring males in high school. In addition, they discovered that differences in the general population became significant in more selective samples. High achievement and highly precocious groups showed the largest differences, again favoring males. Studies by Benbow and Stanley (1980, 1982) specifically addressed this difference at high ability levels. In their research, differences in performance of high achieving seventh and eighth graders on the mathematical portion of the Scholastic Aptitude Test (SAT-M) are much larger for these mathematically talented students than for the general population of SAT-takers. In the highest scoring groups males outnumbered females three to one. In follow-up studies of these same subjects at the high school level, the magnitude of the sex differences in SAT-M performance remained about the same.

Studies frequently conclude that differences on tests are the result of differential performance only on certain types of items. Indeed, the bulk of the research supports a dichotomy of item types: items that are well-defined and resemble material taught in school versus items that are ill-defined and require insightful or unusual solution strategies. Armstrong (1985) and Dossey, Mullis, Lindquist, and Chambers (1988) found that females outperform males on problems where there is an obvious procedural rule, whereas males do better when the problem solving strategy is less clearly defined. This dichotomy can conceivably encompass the categorization of item types defined in other research: the pure mathematics/word problem categorization (McPeek & Wild, 1987; O'Neill, Wild & McPeek, 1989), the algebra/geometry categorization (Boswell, 1985; Doolittle & Cleary, 1987; Hudson, 1986) and the algorithmic/strategic categorization (Doolittle, 1987).

These results support hypotheses put forth by Kimball (1989) in an effort to explain why females tend to get better grades than males in mathematics classes but show poorer performance on mathematics tests. Kimball suggests that sex differences in test performance may be the result of sex differences in the ways students approach and solve problems. Work examining sex differences in the performance of high scorers on the SAT-M also suggests that males and females may use different strategies in solving multiple choice mathematics problems (Gallagher, 1990).

Recent studies examining performance on multiple choice and free-response tests indicate that item format may also affect sex differences in performance. Numerous studies have shown that males perform substantially better than females on the multiple-choice sections of the Advanced Placement tests but that the difference is reduced or favors females in the free-response sections (Breland & Griswold, 1981; Mazzeo, 1987; Peterson & Livingston, 1982).

It is conceivable that some of the differences that were found in these studies may be attributable to the difference in the types of items and scoring that is used on free response items on the Advanced Placement exam. Free response items on this test are fairly complex and generally require several steps to complete. Partial credit is given if some of the steps are completed correctly even if the final response is incorrect. On multiple choice items, on the other hand, credit is only given for a correct final response. It is possible that females' relatively higher performance on free response items on the Advanced Placement exam may result from this type of partial credit scoring.

Performance differences that are related to item format raise questions of whether the same cognitive constructs are assessed by both multiple-choice and free-response item types or whether differences exist in the way males and females approach these item types. To date, conclusions have been ambiguous and tentative at best.

Hogan (1981) concluded from a review of early studies (conducted predominantly in the 1920s and 1930s) that multiple-choice and free-response item types are nearly equivalent. Traub and MacRury (in press), however, discounted most of Hogan's data on the basis of flaws in experimental design or data analysis. Further, Traub and MacRury reviewed additional research conducted since 1970 and concluded that multiple-choice and free-response tests do appear to measure different cognitive abilities, but that the nature of the differences is not clear.

More recently, Traub (in press) synthesized the results of nine studies that examined trait equivalence assessed by multiple-choice and constructed-response tests and concluded that trait equivalence may be dependent on the knowledge domain being assessed. Format effects appeared to be more influential in the writing and word knowledge domains than in the reading comprehension and quantitative domains. However, findings by Bolger and Kellaghan (1990) suggest that multiple choice and free response tests may be measuring different constructs for different groups of examinees.

Bolger and Kellaghan compared the relative performance of males and females on multiple choice and free response items. They report significantly larger differences favoring males on the multiple choice sections of all three domains that they assessed (mathematics, Irish and English achievement), with the largest differences found in mathematics.

There is also disagreement in findings for the direction of the sex differences. Although Bolger and Kellaghan (1990) found that the male advantage was reduced on free-response items, Analyses performed at ETS on pretest data for the New Possibilities test (Schmitt & Crone, 1991) indicated that items in the free-response format were differentially more difficult for females than for males.

These seemingly contradictory findings may be due to differences in study design, type of free-response format, and populations tested. The Bolger and Kellaghan study reported mean differences between males and females. Schmitt and Crone, on the other hand, matched subjects on their total multiple-choice score. A difference between means does not necessarily indicate bias, indeed it may indicate a real difference in ability. However, Bolger and Kellaghan were comparing the magnitude of differences between the same two groups of subjects across two different item formats. In this case, it is not necessary to match on ability since it is the change in the magnitude of the difference that is important, and not the difference itself. As Schmitt and Crone point out, it is possible that matching on total multiple choice score could result in larger differences on the free-response sections.

Another factor which could influence findings is the type of free-response item that was used and the area of mathematics being tested. Males and females may perform differently on free response items that allow for partial credit scoring and grid-in items that are scored either right or wrong. Further, tests of arithmetic may produce different results from tests of higher order problem solving. The Schmitt and Crone study used grid-in items assessing a wide variety of mathematical skills, it is not clear whether the items used in the Bolger and Kellaghan study were grid-in items or items where examinees could receive partial credit.

The work reviewed here indicates that several examinee and test characteristics interact to produce sex differences on tests of mathematics. Sex differences generally appear only in adolescent and post adolescent populations and are most pronounced among high ability groups (Hyde et al., 1990). On multiple-choice tests, males tend to outperform females on items requiring insightful or unusual solution strategies, but the male advantage is rarely found on items requiring the application of algorithms generally taught in school (Gallagher, 1990). Consequently, tests which contain relatively more items requiring unusual solutions, show the largest sex differences. Finally, there is conflicting evidence from the few studies examining sex differences in performance on free-response tests of mathematics (Bolger & Kellaghan, 1990; Schmitt & Crone, 1991). It is possible that this may be due to differences in test content or item type.

The current study seeks to provide preliminary data on how sex and item format relate to the strategies students use to solve mathematics problems. A cognitive analysis was conducted to explore variations in mathematics problem solving on multiple choice and free response items in a small group of subjects. The purpose of this analysis was to provide information that could be used to generate hypotheses about the source of sex differences in performance on multiple choice and free response items.

Method

Twenty-four subjects (12 male and 12 female) were asked to think aloud while solving a set of 13 mathematics problems taken from five disclosed forms of the SAT-M. Six of the 12 males and six of the 12 females received items in their original multiple-choice format, while the other half received items without options (free-response format). Think aloud sessions were audio-taped. According to Ericsson and Simon (1984) think aloud protocols, or concurrent verbal reports, can be considered reliable data on subjects' thought processes. The authors state,

> Our examination of two of the most vigorous challenges to the usefulness of verbal reporting leaves intact our belief that such reports - especially concurrent reports,... of specific cognitive processes - provide powerful means for gaining information about such processes. The concurrent report reveals the sequence of information heeded by the subject without altering the cognitive process ... (p. 30)

Data Collection

Students and their parents were requested to read and sign a consent form with a brief description of the research. The think aloud protocols were collected on audio tape by a female examiner at the subject's school either during school hours or after school. An empty classroom or office was used to keep distractions at a minimum and to insure that audiotapes were intelligible. Problems were presented one at a time, each on a full sheet of paper and subjects were encouraged to use the extra space for scratch. Problem sheets for each student were retained and labeled to aid in transcription of the protocols. Prior to thinking aloud, students were only told that the research they were participating in sought to examine how high scoring students solved problems on the SAT-M. Subjects were asked to read the following written instructions prior to think aloud sessions.

> I am going to give you some math problems from the SAT. I would like you to do these problems in the same way you would if you were taking the SAT except I would like you to "think out-loud" while you solve them. Say everything that you are thinking while you are getting your answer.
>
> You will have four minutes to do each problem. Please use the extra space on each page to write any notes or equations. The first two problems will be for practice. Remember, try to do the problems in exactly the same way you would if you were really taking the SAT. After you finish a problem I will not tell you whether your answer is correct or not, but I may ask you to explain some more about how you got your answer. Do you have any questions?

No other information was given subjects before or during the interview. Interviews were limited to one hour and lasted anywhere from 40 minutes to one hour depending on how much time was available in the student's schedule and how quickly the problems were solved. The examiner intervened only in the following circumstances: a) if the subject paused for more than 30 seconds, the examiner asked "what are you thinking" and; b) if the examiner was unable to understand how the subject solved the problem, the subject was asked "can you tell me how you got that" subsequent to selecting an answer.

Subjects

Subjects were juniors and seniors from public and private high schools in the Princeton, New Jersey area. Eighty-three percent of the subjects were white. This proportion is very close to the 80% reported for high scoring college bound seniors in 1989 (Robertson, 1990). Students were recruited through a mathematics teacher, department head or guidance counselor. The sample consisted of 12 males and 12 females who scored at or above 650 on the SAT-M. This score represents approximately the 95th percentile for females in recent administrations of the test. The 95th percentile for males is approximately 700. Subjects were matched on SAT-M score such that any subject who received multiple-choice items scored within 10 points of another subject of the same sex who received items in the free-response format.

Insert Table 1 about here

Instruments

The Mantel Haenszel Differential Item Functioning procedure (MH DIF) was used to identify items on five disclosed forms of the SAT-M where males and females performed differently. Twenty-one items were identified as favoring males and nine items favored females. Two types of DIF statistics were used to determine differences in item performance; the Mantel Haenszel P-DIF (MH P-DIF) and the Mantel Haenszel D-DIF (MH D-DIF). The P-DIF statistic, based on the normally distributed P+ distribution, is most sensitive to differences in items of middle difficulty where small differences on the ability continuum are equivalent to a larger area under the normal curve than differences at either end of the scale. The D-DIF statistic, on the other hand, is based on delta, which is a linear scale and is, therefore, more sensitive than P-DIF to differences in items of high or low difficulty.

Examinees in the DIF analyses were drawn from one administration of each of the 5 forms that were analyzed: (May '87, November '87, May '88, November '88 and May '89). All analyses were run on subjects who scored at or above 650 which restricted the range of possible scores to 150 points as opposed to the standard 600 point range. This restriction of range in scores was reflected in relatively small DIF values. Consequently a relatively small criterion value was used to identify differentially functioning items. Items were flagged if the MH D-DIF was greater than or equal to .50 and the MH P-DIF was at or above a comparable level for that scale (.05). Although higher levels of the MH D-DIF (1.0 or higher) are used operationally for flagging items on the SAT, a value of .5 or more is considered appropriate for research purposes.

The item set for the think aloud protocols (Appendix A) consisted of 13 items flagged for sex differences in the DIF analysis. Items were selected that could be easily transformed from multiple-choice to free-response format by removing the options. Five of these items favored females and 8 favored males.

The items were pilot tested on four subjects of similar age and mathematical ability to students who would serve as subjects in the protocol analyses. Selection of subjects for the pilot study was based on availability. The four subjects were the first two males and females out of the total pool of subjects who were available for interviews. Results of the pilot testing showed that most students were able to solve items in four minutes or less in the multiple-choice format and that the level of difficulty was appropriate for the group that had been selected. Items were administered in order of difficulty. Two easy items which were not scored were given as practice.

Strategy Classification

Audio tapes were transcribed, and strategies used to solve problems were classified according to the classification scheme developed in Gallagher (1991). The following eight categories were used:

1. **Algorithm:** Solutions that consist **primarily** of computational strategies generally taught in school. This includes computations and algebraic formulas using abstract terms or givens from the problem stem.

2. **Insight with Algorithm:** Solutions that use a mathematical algorithm but are simplified or shortened due to insight, logical reasoning or estimation. This category includes solutions were the student realizes that it is not necessary to complete an equation or algorithm in order to choose an option.

3. **Logic, Estimation or Insight:** Solutions based **primarily** on the application mathematical principals or logic, either alone or in combination with estimation or insight. **These** solutions generally do not employ computations or algorithms, but they may include minor mental calculations. Solutions may require consideration of all cases critical to satisfying given conditions.

4. **Assigns Values to Variables:** Solutions where values are assigned to variables given in the problem stem. This includes "trial and error" solutions using random numbers and solutions where assigned values make operations more "concrete".

5. **Plugs in Options:** Solutions where examinee works backward from options, systematically plugging in choices. This type of strategy can only be used on multiple-choice items.

6. **Guessing:** Solutions based **primarily** on guessing. This includes choices that are based on surface characteristics of options or are unexplained (e.g., "it just looked right"). This category does not include estimations based on partial solutions.

7. **No Strategy:** Examinee did not formulate a solution strategy. This category applies only to instances where no attempt at solution was made (e.g., "I don't know how to do this"). It does **NOT** include items where the examinee attempted but failed to solve.

8. **Misinterpretation:** This category applies only to solutions where the examinee clearly misread the item stem or diagram or where the stem or diagram were misinterpreted. It does not include faulty solutions caused by misunderstanding or misuse of mathematical terms and concepts or unfinished solutions.

Examples of solutions falling in these categories are provided in Appendix B.

Solutions to each problem were coded by one rater using the eight categories listed above. A second rater was then trained by the first rater to use the coding system. Training consisted of a review of the categories followed by a discussion of prototypical solutions to several problems. The second rater then "practiced" coding solutions on two problems. Ratings for solutions to these problems were then reviewed and discussed with the first rater. Finally, the second rater coded solutions to the remainder of the 13 problems (263 solutions). The two problems used for practice coding were not included in the analyses. Coding by both raters was "blind" in that coders did not know the subject's sex. Initially, there was 80% agreement between the first and second raters. Subsequent to discussion of discrepant ratings, agreement was 100%.

Results

Although subjects were matched on SAT-M score within sex, it was not possible to match subjects precisely by sex. Table 1 displays the distribution of SAT-M scores by sex and format of items they received. Because score distributions for males and females were somewhat different, analyses were performed only within sex so that ability and effects due to item format would not be confounded. In addition, large differences in performance and strategy use were rarely found on any individual item. It is likely that this was due to the small sample size. Therefore, all analyses were run across items.

Responses to items in multiple-choice and free-response formats were examined to determine if there were differences within sex in the proportion of items answered correctly. The Mantel Haenszel chi square statistic was used to examine differences across items. There was a significant difference between multiple-choice and free-response performance for females $X^2(1) = 8.4$, $p<.01$, with females who received the multiple-choice items performing substantially better than those who received free-response items.

Typically, errors for females in the free-response format consisted of calculation errors or errors in setting up a problem. In the multiple-choice format, calculation errors virtually disappeared, and errors for females consisted mainly of difficulty in setting up problems or errors caused by their failure to complete a problem. For example, on problem number two, where a correct solution requires solving for a variable N, and then using that value to find the length of line segment PR. Several students solved for N but failed to continue solving for the length of PR.

Performance differences for males on the two item types were in the same direction as those found for females, however, this difference was not significant. Errors committed by males also followed patterns that were similar to those found for females.

To allow for larger cell sizes in the analyses, strategy categories were grouped into strategies that apply standard algorithms and those that require insight. Strategy types 1 (applies a computational algorithm), 4 (substitutes values for variables) and 5 (works backward from options) could all be grouped in an "algorithm" category since they all are fairly standard computational strategies. Strategies 2 (insight with algorithm) and 3 (logic, estimation or insight) could be grouped in an "insight" category since they both require insight based on mathematical principles.

These collapsed strategy categories were examined to determine whether they varied by item format. The difference in strategies used on multiple-choice and free-response items was only significant for males MH $X^2(1) = 8.22$, $p<.01$. On multiple-choice items strategies used by males were almost equally divided between algorithmic and insightful strategies (algorithmic strategies were us d 47% of the time). On free-response items, however, males were more likely to use algorithmic rather than insightful strategies (algorithmic strategies were used 66% of the time). Appendix B provides examples of algorithmic and insightful solutions. Although there was no difference in strategies used by females on the two item types, it should be noted that the majority of solutions generated by females were with algorithmic strategies. These strategies made up 85% of strategies used on multiple-choice items and 73% of those used on free-response items.

In addition to the eight strategy categories used in coding, transcripts were examined for solution characteristics that might be related to the free-response format. Specifically, solutions were examined in the multiple-choice format to determine whether students initially generated an answer that was not an option, and the extent to which they reworked their solutions to get a correct answer. Free-response solutions were examined to determine how frequently students generated answers that were not options in the multiple-choice version of the items. There was no significant difference by item format for either sex in the frequency of answers generated that were not options in the multiple-choice version of an item.

However, when answers that were corrected were removed from the analysis, there was a significant difference in the mean frequency by format. For both males and females the frequency of wrong answers produced when the initial response was not an option was significantly greater in the free-response format than in the multiple choice format ($t(12) = 3.55$ $\underline{p} < .005$ for females and $t(12) = 2.6$, $\underline{p} < .025$ for males). Effect sizes for these differences were calculated, and it was apparent that although both showed large differences, the largest difference between the means was for females ($\underline{d} = 2.07$ for females and $\underline{d} = 1.5$ for males).

Finally, the time it took students to solve multiple-choice and free-response items was examined. Analyses showed no significant difference in time by format for either sex. It is possible that the "think aloud" format of interviews may have affected the time it took students to solve problems. It is conceivable that the multiple choice items may have been solved more quickly without this requirement.

Discussion

The study presented here has attempted to provide preliminary data from which hypotheses could be generated about how sex and item format are related to the strategies high ability students use to solve mathematics problems. Work by Hyde et al. (1990) indicates that performance differences are substantially larger in high ability groups. This suggests that different factors may be responsible for sex differences in mathematics performance at various levels of the ability continuum. For this reason, results of this study may not generalize to other levels.

It should also be noted that males and females in this study were not precisely matched on any ability measure. The differences in the distribution of SAT-M scores for these subjects is relatively small, with substantial overlap between the two groups, however, these differences may have affected the outcomes. Further work that employs precise matching of males and females on an ability measure will be necessary before we can rule out hypotheses that attribute sex differences in solution approach to ability differences.

The data analysis indicated that both males and females performed better in the multiple-choice format than in the free-response format. However, the performance difference was somewhat greater for females than for males. This agrees with other findings for SAT-M items in the two formats (Schmitt & Crone, 1991).

In addition, females in this study tended to rely more heavily on algorithmic strategies in both the multiple-choice and free-response formats. Males, on the other hand, varied their strategy according to item format. They were generally less likely than females to use algorithmic strategies, but appeared to use them more often on free-response items than on multiple-choice items.

For example, on problem number 5 in Appendix A, students are asked how much of a particular type of coffee is required in 50 pounds of a blend of coffee costing $5. There are at least three ways to solve this problem; using a computational strategy, or using one of two logical strategies. The computational strategy is fairly complex and requires some time and thought to set up an equation, indeed, many who chose this strategy failed to set up an equation that worked. All of the females in the study used some type of computational strategy to solve this problem, regardless of the item format.

Two logical strategies were used by males in order to solve this problem. One strategy can only be used in the multiple-choice version since it relies on the distractors. Using this strategy, one can see that more than half of the mixture must be the cheaper coffee (espresso) since the price per pound for the mixture is less than half the sum of the price per pound of each type. Using this information, there is only one possible option that could be correct since all of the others are greater than 25 (or half of the 50 pounds of the mixture).

The other type of logical strategy which was used by males on both multiple-choice and free-response versions of the item pursues the following logic. The cost of the blend, ($5.00) is not equidistant between the cost of the two components ($3.00 and $8.00), but it is close because equal amounts of the components would produce a blend that costs $5.50. The price of one component is three units away from the cost of the blend, and the price of the other is two units away. Therefore, the blend consists of components in a ratio of three to two. There must be more of the cheaper component, since the blend costs less than $5.50, so the blend must be made up of three parts cheap to two parts expensive.

On this problem then, students (males) who tended to look for short cuts using their knowledge of mathematical principles avoided the pitfalls of extensive computation. Even if the student who used an algorithmic strategy was able to set up an equation that made sense, there is still room for calculation errors. Although these errors will be obvious in multiple-choice versions of the item where the generated answer would not be one of the options, they will go undetected in the free-response version.

The examination of responses generated that were not options supports Bridgeman's notion that students use the options in multiple choice items as hints. Although there was no difference for either sex in the number of "non-option" responses generated in multiple-choice and free-response formats, females were able to detect and correct proportionally more of their errors in the multiple-choice format than were males. This could be the result of two factors. First, females were more likely to use algorithmic strategies, which are easier to rework in order to detect calculation errors, than errors based on faulty logic. Indeed, the analyses of errors showed that when males generated an answer that was not an option, they were more likely than females to guess, and the guesses, were generally incorrect. In 75% of the cases where males guessed, the chosen option was not correct.

An example of solution strategies used in multiple-choice and free-response items serves to illustrate the types of strategy differences found for males. As noted earlier, there was no clear delineation of strategy types within any individual item, however, strategy differences on this particular item provide an example of the type of differences found when strategy use across all items was examined.

Item number 9 presents a situation where the student is asked to determine how much Mrs. Smith spent on lunch if she subsequently spent a certain percentage, and was left with $5.00. The algorithmic solution to this problem creates an equation with variables in the order that is given in the problem (e.g., $X-(.2)X-(.75)(.8X)=\$5.00$). This solution requires some fairly complex computations. However, if the solution applies the informaiton in the reverse order, then only fairly simple computations are necessary. This more insightful solution proceeds in the following manner: if 5 equals 25%, or 1/4 of what was left after lunch, then 4x5=20 which is what she had after lunch, so 20 is 80% or 4/5 of what she originally had, divide, 20 by 4 to get 5 and add it to get 25, which is what she had before lunch, so she spent $5 on lunch.

The sex differences in solution strategies that were found here provide insight into the discrepancy between observed patterns of sex differences in mathematics test scores and class grades discussed by Kimball (1989). If classroom measures require the application of solution strategies that have been taught (familiar, well-defined strategies), then students who are more apt to use these strategies spontaneously will outperform students who are less likely to use them. On the other hand, standardized tests that contain items whose solutions have not been explicitly taught would yield sex differences favoring males who are more apt to use more creative solution strategies.

These strategy differences may also be related to differences found in crystallized (G_c) and fluid intelligence (G_f) (Snow 1980, 1982). There is also evidence to suggest that females tend to do better on tasks that reflect G_c, which are generally more familiar in content and format, whereas males tend to do better on G_f tasks that present novel or unfamiliar problems (Mandinach, 1984; Mandinach & Corno, 1985). Furthermore, the ability to know when to change strategies has been found to be related to G_c - G_f differences (Bethell-Fox, Lohman, & Snow, 1984; Kyllonen, Lohman, & Woltz, 1984) and gender (Mandinach, 1984).

From these results, one can construct a preliminary sketch of differences in how males and females approach multiple-choice and free-response mathematics items. Findings indicating that males are more likely to use insightful strategies than females and that they use them to a greater extent on multiple-choice items than on free-response items, suggest that males may prefer this type of strategy to more algorithmic strategies. It seems reasonable that they might use them more often on multiple-choice items where the options help them to determine whether their response is correct or where they can estimate using the options.

Females in this analysis appear to prefer more algorithmic strategies regardless of the item format. It is possible that the larger difference between the mean frequencies of "non-option" answers found for females is related to the fact that they used algorithmic strategies more frequently. Algorithmic strategies rely heavily on computation, and are often complex, allowing greater opportunity for careless errors than solutions that are based on insights born of a knowledge of how mathematics works as a system. Perhaps one of the reasons that the gap between multiple-choice and free-response performance is larger for females is that in the multiple-choice format, options can be used as a guide to detecting calculation errors or other careless errors. In this small sample, analyses suggested that this could well be the case.

Future Research

Although findings of this study offer a reasonable hypothesis for why free-response items on the SAT-M may be differentially more difficult for high scoring females than for their male counterparts, the small size of the sample used and the selectivity of the sample dictates that these findings be treated as preliminary and only for use in generating hypotheses. Variations on the current study need to be performed at other ability levels to determine how strategy differences vary by the ability level of the examinees. Once this has been accomplished, the hypotheses generated by these small scale studies can be tested in more traditional, large scale studies.

For example, to test the hypotheses generated by the current study, a test could be constructed that contains two types of items: a) items that can only be solved using algorithmic or computational strategies, and b) items that can only be solved using insightful strategies that require little or no computation. These two item types could then be administered in one of three formats: a) multiple choice items with distractors composed of the most commonly found wrong answers that were given when the item is administered in the free response format (pilot testing would be necessary to obtain these distractors), b) multiple choice items with distractors generated under the current system, and c) free response items. These items could then be administered to large groups of subjects who were carefully matched on mathematics ability and background variables such as course-work in mathematics, attitudes towards mathematics and test anxiety. In this way it could be determined whether strategy differences found in the current study exist in the larger population of test takers.

REFERENCES

Armstrong, J.M. (1985). A national assessment of participation and achievement of women in mathematics. In S.F. Chipman, L.R. Brush, & D.M. Wilson (Eds.), Women and mathematics: Balancing the equation (pp. 59-94). Hillsdale, NJ: Lawrence Erlbaum Associates.

Benbow, C.P., & Stanley, J.C. (1980). Sex differences in mathematical ability:Fact or artifact? Science, 210, 1262-1264.

Benbow, C.P., & Stanley, J.C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. American Educational Research Journal, 19(4) 598-622.

Bethell-Fox, C.E., Lohman, D.T., & Snow, R.E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. Intelligence, 8, 205-238.

Bolger, N. and Kellaghan, T, (1990). Method of measurement and gender differences in scholastic achievement. Journal of Educational Measurement, 27(2), 165-174.

Breland, H.M. and Griswold, P.A. (1981). Group comparisons for basic skills measures. College Board Report No. 81-6, New York, NY: College Entrance Examination Board.

Doolittle, A.E. (1987). Gender differences in performance on mathematics achievement items. ACT Research Report, 87-16.

Doolittle, A.E., & Cleary, T.A. (1987). Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24, 157-166.

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M., & Chambers, D.L. (1988) The mathematics report card: Are we measuring up? Trends and achievement based on the 1986 National Assessment. Princeton, NJ: The Nation's Report Card, NAEP, Educational Testing Service.

Feingold, A. (1988). Cognitive gender differences are disappearing. American Psychologist, 43(2), 95-103.

Gallagher, A.M. (1990). Sex differences in the performance of high-scoring examinees on the SAT-M. College Board Report No. 90-3, New York, NY: College Entrance Examination Board.

Gallagher, A.M. (1991). Sex differences in problem solving strategies used by high scoring examinees on the SAT-M. Unpublished doctoral dissertation, Rutgers University, NJ.

Hogan, T.P. (1981). Relationship between free-response and choice-type tests of achievement: A review of the literature. Green Bay, Wisconsin: University of Wisconsin. (ERIC Document Ed 224 811)

Hudson, L. (1986). Item-level analysis of sex differences in mathematics achievement test performance. Dissertation Abstracts International, 47(2) Order no. DA8607283.

Hyde, J.S. (1981). How large are cognitive gender differences? American Psychologist, 36(8), 892-901.

Hyde, J.S., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematical performance: A meta-analysis. Psychological Bulletin, 107(2), 139-155.

Kimball, M.M. (1989). A new perspective on women's math achievement. Psychological Bulletin, 105, 198-214.

Kyllonen, P.C., Lohman, D.F., & Woltz, D.J. (1984). Componential modeling of alternative strategies for performing spatial tasks. Journal of Educational Psychology, 76, 1325-1345.

Maccoby, E.E., & Jacklin, C.N. (1974). The psychology of sex differences. Stanford: Stanford University Press.

Mandinach, E.B. (1984). The role of strategic planning and self-regulation in learning an intellectual computer game. Unpublished doctoral dissertation, Stanford University, CA.

Mandinach, E.B. & Corno, L. (1985). Cognitive engagement variations among students of different ability level and sex in a computer problem solving game. Sex Roles, 13, 241-251.

Mazzeo, J. (1987). "Summary statistics for males and females on multiple-choice and free-response sections of AP exams". Unpublished memorandum issued September 22, 1987.

McPeek, W.M., & Wild, C.L. (1987). Characteristics of quantitative items that function differently for men and women. Paper presented at the annual meeting of the American Psychological Association, New York.

O'Neill, K., Wild, C.L,. & McPeek, W.M. (1989). Gender-related differential item performance on graduate admissions tests. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA

Peterson, N. & Livingston, S.L. (1982). English Composition Test with Essay: A descriptive study of the relationship between essay and objective scores by ethnic group and sex. ETS Statistical Report No. SR-82-96.

Robertson, N. (1990). 1989 College-bound seniors high-scoring on SAT-V (>600) and on SAT-M (>650), with contrast to total group of 1989 college-bound seniors. Unpublished manuscript: Educational Testing Service.

Schmitt, A.P. & Crone, C.R. (1991). Alternative mathematics item types: DIF issues. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, Illinois.

Snow, R.E. (1980). Aptitude processes. in R.E. Snow, P-A. Federico, & W.E. Montague (Eds.) Aptitude learning and instruction. vol. 1. Hillsdale, NJ: Earlbaum.

Snow, R.E. (1982). Training of intellectual aptitude. in D.K. Detterman, & R.J. Sternberg (Eds.) How and how much can intelligence be increased. Norwood, NJ: Ablex.

Traub, R.E. (in press). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R.E. Bennett & W.C. Ward, (Eds.), Construction versus choice in cognitive measurement. Hillsdale, NJ: Earlbaum.

Traub, R.E., & MacRury, K. (In press), Multiple-choice vs. free-response in the testing of scholastic achievement. In K. Ingenkamp & R.S. Jager (Eds.), Tests und trends 8: Jahrbuch der padagogischen diagnostik. Weinhaim und Basel, Germany: Beltz Publishing Company.

Table 1

Distribution of SAT-M Scores by Sex

Format	Score	Males	Females
Multiple-Choice	680	1	1
	690	0	0
	700	0	0
	710	0	1
	720	1	1
	740	0	2
	750	0	1
	760	1	0
	780	1	0
	790	1	0
	800	1	0
Free-Response	680	1	0
	690	0	1
	710	0	2
	720	1	0
	730	0	1
	750	0	2
	770	2	0
	800	2	0

APPENDIX A

Multiple-Choice and Free-Response Versions of

Items Used in Protocols

1. Mary's present age is half Paul's age. In 4 years, Paul will be n years old. In terms of n, how old is Mary now?

(A) $\frac{n}{2} - 2$

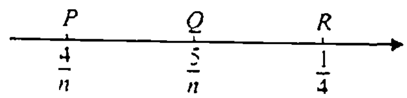
(B) $\frac{n}{2}$

(C) $\frac{n}{2} + 2$

(D) $\frac{n}{2} + 4$

(E) $2n + 4$

1. Mary's present age is half Paul's age. In 4 years, Paul will be n years old. In terms of n, how old is Mary now?

2. 

If $PQ = QR$ on the number line above, what is the length of PR?

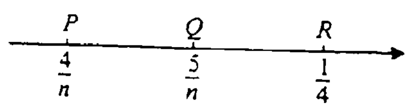
(A) $\frac{1}{12}$

(B) $\frac{1}{9}$

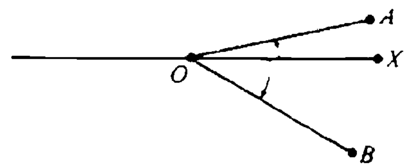
(C) $\frac{1}{8}$

(D) $\frac{1}{6}$

(E) $\frac{3}{16}$

2. 

If $PQ = QR$ on the number line above, what is the length of PR?

3. 

Segments OA and OB, shown in the figure above, start on OX at the same time, and revolve simultaneously in the plane in opposite directions about point O. If OA revolves at 2° per second and OB revolves at 6° per second, in how many seconds after they start will OA and OB first meet?

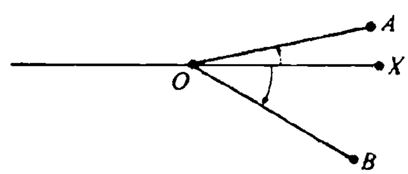
(A) $22\frac{1}{2}$

(B) 45

(C) 50

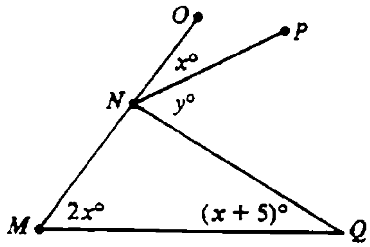
(D) 90

(E) It cannot be determined from the information given.

3. 

Segments OA and OB, shown in the figure above, start on OX at the same time, and revolve simultaneously in the plane in opposite directions about point O. If OA revolves at 2° per second and OB revolves at 6° per second, in how many seconds after they start will OA and OB first meet?

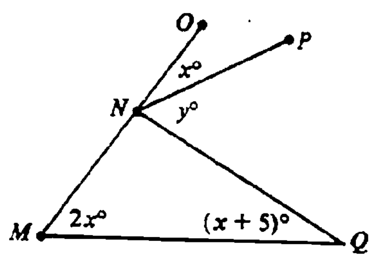
4.



In the figure above, N lies on line segment MO. Which of the following gives y in terms of x?

(A) $2x$
(B) $2x + 5$
(C) $3x + 5$
(D) $90 - x$
(E) $180 - 3x$

4.

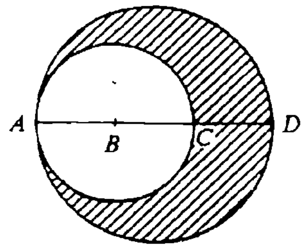


In the figure above, N lies on line segment MO. Which of the following gives y in terms of x?

5. A blend of coffee is made by mixing Colombian coffee at $8 a pound with espresso coffee at $3 a pound. If the blend is worth $5 a pound, how many pounds of the Colombian coffee are needed to make 50 pounds of the blend?

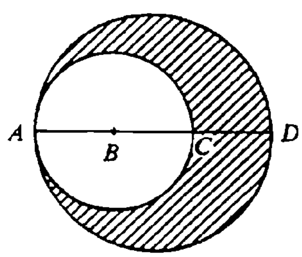
(A) 20
(B) 25
(C) 30
(D) 35
(E) 40

5. A blend of coffee is made by mixing Colombian coffee at $8 a pound with espresso coffee at $3 a pound. If the blend is worth $5 a pound, how many pounds of the Colombian coffee are needed to make 50 pounds of the blend?



6. In the figure above, AC and AD are diameters of the small and large circles, respectively. If $AB = BC = CD$, what is the ratio of the area of the shaded region to the area of the smaller circle?

(A) 1 : 1
(B) 3 : 2
(C) 4 : 3
(D) 5 : 4
(E) 9 : 4



6. In the figure above, AC and AD are diameters of the small and large circles, respectively. If $AB = BC = CD$, what is the ratio of the area of the shaded region to the area of the smaller circle?

7. An airplane traveled a distance d in t hours, where $t > 1$, and arrived one hour late. The airplane would have arrived on time if it had traveled at what rate per hour?

(A) $t - 1$

(B) $\frac{d}{t} - 1$

(C) $\frac{d}{t} + 1$

(D) $\frac{d}{t - 1}$

(E) $\frac{d}{t + 1}$

7. An airplane traveled a distance d in t hours, where $t > 1$, and arrived one hour late. The airplane would have arrived on time if it had traveled at what rate per hour?

(A) $t - 1$

8. If the average (arithmetic mean) of six numbers is -6, and the sum of four of the numbers is 20, what is the average of the other two numbers?

(A) 7
(B) 8
(C) -8
(D) -28
(E) -32

8. If the average (arithmetic mean) of six numbers is -6, and the sum of four of the numbers is 20, what is the average of the other two numbers?

9. Ms. Smith spent 20 percent of the money she had in her purse on lunch and 75 percent of what was left after lunch on groceries. If she then had $5.00 left, how much had she spent on lunch?

(A) $10.00
(B) $8.50
(C) $6.00
(D) $5.00
(E) $4.50

9. Ms. Smith spent 20 percent of the money she had in her purse on lunch and 75 percent of what was left after lunch on groceries. If she then had $5.00 left, how much had she spent on lunch?

10. Each plant of a certain variety yields 50 seeds in the early fall and then dies. Only 40 percent of these seeds produce plants the following summer and the remainder never produce plants. At this rate. a single plant yielding seeds in 1986 will produce how many plants as descendants in 1989 ?

(A) 60
(B) 400
(C) 8.000
(D) 16.000
(E) 32.000

10. Each plant of a certain variety yields 50 seeds in the early fall and then dies. Only 40 percent of these seeds produce plants the following summer and the remainder never produce plants. At this rate. a single plant yielding seeds in 1986 will produce how many plants as descendants in 1989 ?

11. x increased by 10% of x yields y.
y decreased by 50% of y yields z.
z increased by 40% of z yields w.

According to the statements above, w is what percent of x ?

(A) 10%
(B) 33%
(C) 77%
(D) 81%
(E) 100%

11. x increased by 10% of x yields y.
y decreased by 50% of y yields z.
z increased by 40% of z yields w.

According to the statements above, w is what percent of x ?

12. If n oranges cost p dollars, at this rate how many dollars will 5 oranges cost, in terms of n and p?

(A) $5np$

(B) $\frac{n}{5p}$

(C) $\frac{p}{5n}$

(D) $\frac{5n}{p}$

(E) $\frac{5p}{n}$

12. If n oranges cost p dollars, at this rate how many dollars will 5 oranges cost, in terms of n and p?

13. The average (arithmetic mean) of 5 integers is greater than 27. If the average of the first 4 integers is 22, what is the least possible value of the 5th integer?

(A) 32
(B) 33
(C) 47
(D) 48
(E) 49

13. The average (arithmetic mean) of 5 integers is greater than 27. If the average of the first 4 integers is 22, what is the least possible value of the 5th integer?

APPENDIX B

Examples of Solution Strategies

Problem #12
Classified as category 1 (algorithmic).

According to the statements above, W is what percent of X? X increased by 10% of X yields Y, (mumble) Y decreased by 50% of Y yields Z. Z increased by 40% of Z yields W. They want what... W is what percent of X? Um, so let's say X times 1.1 equals Y, Y decreased by 50% equals Z, so then X times 1.1 over 2 equals Z and Z increased by 40% yields W. So, X times 1.1 percent 1.1 percent over 2, all times 1.4 should equals W. So then X is let's see, um, (mumble) 1.4 get rid of that. Times 2, .7, get rid of that and then divide by 1.1 would be C, 77 and the decimals moves in 2 places .77, so W is .77 equals X, so let's see I guess it would be C. Oh wait, W is what percent of X, yeah, I would say 77%.

Problem #2
Classified as category 1 (algorithmic).

Um, I'm, one-fourth minus 4 over N to get the length of PR, um (pause) and then um 5 over N minus 4 over N equals one-fourth minus 5 over N. 5 minus 4 equals one-fourth N minus 5, so 10 minus 4 equals one-fourth N, 6 equals one-fourth N, N equals 24. And then one-fourth minus 4 24ths equals PR, then 6 over 24 minus 4 over 24 equals PR, um then 2 over 24 equals PR, so 1 over 12 equals PR.

Problem #2
Classified category 2 (insightful).

PQ equals QR on the number line above, what is the length of PR? 1 quarter, um let's see. (Pause) uh, I guess you would just multiply 1 quarter times 6, so that you'd get the logical progression here from 4 to 5 to 6. That would be 6 over 24 and then these would have to have the same denominator so it would be 5 over 24 and 4 over 24, so 6 over 24 minus 4 over 24 would be 2 over 24, which would equals one-twelfth which would be A.

Problem #5
Solution classified category 2 (insightful).

(Pause) (mumble) first you'd make a ratio, of some sort. Seeing that uh, from 3 to 2, from 3 to 5 is, is, is a quantity of 2 and from 5 to 8 is 3, you can right away say that there is gonna be more of the espresso coffee in it because, it's a little, it's less than half. And the midpoint would be 5.5 actually. So um, to make the ratio you'd probably it would be uh, looks like a 40, 60 ratio, which 60% of it would be espresso and 40% of it would be the Columbian coffee. So to make 50 pounds of the blend you would uh, 40% of it would have to be the Columbian, so therefore 40% of 50 is (pause) would be 20. So I'm gonna A, is 20 pounds and then actually I'd check it by saying uh, if you put 20 dollars, 20 pounds of it at $8, would be 160, would give you 160 dollars. And uh, you take 60% of it which would actually be uh, very (mumble) 30 pounds, you times it by the 3 and you get 90. And when you total them you get 250 dollars and $250 equals 5 times, 5 pounds times the 50 pounds of the blend, so it checks out.

Problem #3
Classified as category 3 (insightful).

Let's see (pause) segments OA (pause) ok OA revolves 2 degrees per second so I'll just write that down, per second. And then OB revolves 6 degrees per second, so I'll write that down. And so the, like, they'll be, when they meet, it will be after a certain amount of time, the time will be constant since they start from um, the same point, the same time. So, certain time times 2, 2 degrees per second is equals the same time times 6 degrees per second (pause) (mumble) or this moves three times as fast as this, so the ratio would be 3 to 1. In which case that would mean, that um, if you have a circle and you divide it into 4 parts and you get 1, 2, 3, 8 over here. They will be, they will meet after uh, so if, if they meet at 90 degrees (pause) then just divide 90 degrees by 2 and 2 degrees per second you get 45 seconds.

Problem #10
Classified as category 3 (insightful).

Each plant of a certain variety yields 50 seeds in the early fall and dies. Only some of these seeds produce plants (mumble) 40, 40 times 50 um, at this rate a single plant yielding seeds in 1986 will produce how many plants as descendants in 1989? Uh, let's see not many made it originally so, four-tenths of that is uh, eight-fifths, uh (mumble) 50 times .4 (mumble) 20 I guess (mumble) 20 and (mumble) will produce how many plants as descendants in 1989? (Mumble) so in 87, there are only 20 plants, (mumble) how many plants as descendants in 89? Well, I d n't know, if they died I wouldn't think there would be that many, but uh (pause) I guess it would be 20 (pause) to the third, so it would be 20, 400, I guess it would be 8000 because I guess if that's (mumble) time in years.

(Mumble) I guess it would be 8000, but if they died it seems to be that the last one would get small, fewer I don't know.

Problem #12
Classified as category 4 (algorithmic).

If N oranges cost P dollars, at this rate how many dollars (pause) ok I would plug in numbers for again probably if, if just say 10 oranges cost um, (pause) (mumble) (pause) uh so if 20 oranges costs um, 10 dollars, at this rate how many dollars will 5 oranges cost, in terms of N and P? Well (pause) you have to go (pause) the amount, the money divided by how many to find out how much each one costs. Which would show you 50%, 50 cents in this case, and times th by 5 (pause) so that would at least work. 50 and if 10, I would just check I don't (pause) 10 oranges, 10 dollars, 5, yeah so it would have to be, the amount that it costs which would be P, over the amount that there is to find out how much each one costs and then you'd have to times it by 5, to get the answer.

Problem #8
Classified as category 4 (algorithmic).

I would plug in numbers that would make it work and then, so, let's say an airplane traveled a distance of (pause) this is gonna (pause) 240 miles in (pause) oh wait let's say 360 miles in (pause) 7 hours. Cause it, (mumble) one hour late, where, so 60 miles per hour would go, airplane would have arrived on, arrived on time if it had traveled (pause) oh. (Pause) you go D (pause) and the T hours is one hour late, so it would be adding an extra hour, so you'd have to find out what rate they should have traveled, to get them in hour earlier. Airplane would have arrived on time if it had traveled, arrived one hour late. So you go T minus 1, cause if they did that 360 miles in 7 hours, you'd want to figure out how, long it would take them in 6 hours, how many miles they'd have to do in 6 hours to get there on time. So you'd have to go T minus 1 um, (pause) and the distance, you'd have to divide the distance, distance equals V times T and T is the time, and so to find the velocity or rate, you'd have to go D over T minus 1.

Problem #3
Classified as category 5 (algorithmic).

(Mumble) (pause) how many seconds (mumble) first meet. (Pause) and revolve simultaneously in a plane (pause) this is not a counting one. So you want to find a multiplier that's gonna make it add up to, I guess 360. See I would just guess at answers, or look at their answers. 50 times 6 degrees would be 300 degrees, would be 400 degrees that's too much. Uh, 22 and a half sounds good , that would be (mumble) 32, 135 that would put it around here somewhere. hm... 2 degrees would be 45 (mumble) so they're not going to meet. So I guess 45 is double. 6 times 45 is 270, and 90 would be 360. 45.

<u>Problem #5</u>
Classified as category 5 (algorithmic).

(Pause) alright the whole thing, it's $5 a pound and there's 50 pounds of the whole thing, costs uh, wait I was going to say $10 that's wrong. $250, that's not going to help me out but...(pause) i'm just gonna plug back the answers to see if I can do it backwards. 20 pounds, would be 160, you can do it that way. multiply 20 times 8, it's 160 and then you have to multiply 30 times 3 it's 90, (mumble) wait, does that make sense?... Then you have, you have 20 times (mumble) Yeah.

<u>Problem #5</u>
Classified as category 6 (guessing).

(Pause) This is one of those I start guessing at N's until I figured it out. So um (mumble) 4 fifths to 5 fifths (mumble) hum, 1 quarter is gonna be (mumble) hum. (Pause) (mumble) let me try a multiple of 4 if they're going to be evenly spaced so 8, 12, be a third, 6 12ths that would be a half. Then 16ths would be 4 16ths, 6 16ths not equivalent. Hum, hum, 3 thirty-seconds. Hum, I don't know. I'd guess cause I always do. I'll guess 1 sixth. I just, I always guess, I know you're not suppose to but.

<u>Problem #4</u>
Classified as category 6 (guessing).

In the figure above, N lies on segment MO. Which of the following gives Y in terms of X? Uh, (pause) Y is gonna be equals let's see, do I have any parallel's here? No I don't, uh, got the midpoint there, ok i'm gonna want to find out what this, it's not a right triangle so I can't figure out what, what MNQ is. Uh, oh I know that uh 180 equals uh X plus Y plus whatever that uh, what MNQ is. Uh, midpoint, midpoint, I don't know that uh, MNQ is a 90 degree angle uh, so I can't use 180. But if it's a straight, it is a straight line, but I don't know the third value. Uh, N lies on the segment MO. They really don't give me too many values here. Uh, Y is gonna be equals let's just say that I connect P and Q to make another triangle, uh I don't know that anything is parallel or perpendicular, doesn't appear to be. Uh, (pause) hum, if I assume, I know it has something to do with one or the other values. If I assume that, that ONQ is a 90 degree angle, that would make Y equals 90 minus X, but I don't know that for a fact because I, I only know that it's a point, I don't know that it's a perpendicular. 3X, it might have something to do with the X plus 5, uh (pause) hum. I'd pick D the 90 minus X for uh (pause), because I have to pick one.

Problem #12
Classified as category 7 (no strategy).

N oranges, P dollars, at this rate how many dollars will 5 oranges cost? N oranges cost P dollars, alright so you get N oranges, N oranges, for P dollars. Alright, at this rate how many dollars will 5 oranges cost?
(Pause) alright so uh, N oranges, so that's any number of oranges cost P dollars, uh (pause) that means uh, how many dollars will 5 oranges cost? N oranges cost P dollars, (pause) um, so I guess you're gonna have to uh probably find out I guess change the N to 5 or something. Um, that means 5 oranges (pause) would be uh, is equals uh, (pause) at this rate, if N oranges cost P dollars, if any number of oranges. Um, how many dollars will 5 oranges cost, in terms of N and P? N oranges cost P dollars (pause) that means uh, (pause) wow N oranges cost P dollars, and 5 oranges. Oh boy, (pause) that means 5 oranges (mumble) P dollars, (mumble) times a certain amount. Then if you, if you say that 5 oranges would then cost alright so that means that at 5N it would be, N, if N equals 5, that means N would equals 5, and if N would equals P (pause) so it could be. So that means N equals 5 (pause). Alright, I don't know this one.

Problem #3
Classified as category 8 (misreads or misinterprets problem).

(Pause) alright um, (pause) I don't think you can determine it cause I think you're gonna need to know how big the angles are, to know how far they have to, how many, how far they have to move. I think they have to tell you (mumble).